

Cluster Analysis: nagdmc_cind

Purpose

nagdmc_cind computes cluster indicators following an hierarchical clustering.

Declaration

```
#include <nagdmc.h>
void nagdmc_cind(long nrec, double mergedist[], long denord[], double dendist[],
                 long k, long ic[], long nic[], long *nclust, int *info);
```

Parameters

- 1: **nrec** – long *Input*
On entry: the number of consecutive records, beginning at **rec1**, used in the analysis.
Constraint: **nrec** > 1.
- 2: **mergedist**[nrec-1] – double *Input*
On entry: the distance at which clusters are merged as returned by **nagdmc_hclust**.
- 3: **denord**[nrec] – long *Input*
On entry: the order of the data records required for plotting a dendrogram as returned by **nagdmc_hclust**.
- 4: **dendist**[nrec] – double *Input*
On entry: **dendist**[*i*] is the distance at which cluster **denord**[*i*] merges with cluster **denord**[*i* + 1] as returned by **nagdmc_hclust**. (**dendist**[nrec – 1] contains the maximum distance.)
- 5: **k** – long *Input*
On entry: the number of clusters to form.
Constraint: 1 < **k** < **nrec**.
- 6: **ic**[nrec] – long *Output*
On exit: **ic**[*i*] contains the allocation of the *i*th record to one of the **k** clusters numbered: 0, 1, ..., **k** – 1, for *i* = 0, 1, ..., **nrec** – 1.
- 7: **nic**[k] – long *Output*
On exit: **nic**[*i*] contains the number of data records belonging to the *i*th cluster formed, for *i* = 0, 1, ..., **k** – 1.
- 8: **nclust** – long * *Output*
On exit: the number of clusters returned which may be less than the value of **k** due to ties.
- 9: **info** – int * *Output*
On exit: **info** gives information on the success of the function call:
 - 1: due to ties, there are fewer than **k** clusters..
 - 0: the function successfully completed its task.
 - i*; *i* = 1, 5: the specification of the *i*th formal parameter was incorrect.
 - 54: the information from **nagdmc_hclust** has become corrupted.
 - 100: an internal error occurred during the execution of the function.

Notation

nrec the number of data records, *n*.
k the number of clusters, *k*.

Description

Cluster analysis aims to group n data records into a number of more or less homogeneous groups or clusters. With agglomerative clustering methods, such as [nagdmc_hclust](#), an hierarchical tree is determined by starting with n clusters each with a single (unique) data record and then at each of $n - 1$ stages merging two clusters to form a larger cluster until all data records belong to a single cluster.

nagdmc_cind takes the information from the hierarchical clustering and allocates data records to a given number of clusters. However, it is not always possible to compute the number of clusters requested due to ties in the distance matrix.

If there are k clusters then the indicator variable will assign a value between 0 and $k - 1$ to each data record to indicate to which cluster it belongs. Data record 0 always belongs to cluster number 0.

References and Further Reading

Everitt B S (1974) *Cluster Analysis* Heinemann.

Krzanowski W J (1990) *Principles of Multivariate Analysis* Oxford University Press.

See Also

nagdmc_hclust	computes an hierarchical clustering.
nagdmc_tab2	cross-tabulates known groupings and cluster memberships.
hclust_ex.c	the example calling program.